Preprints of the 19th World Congress
The International Federation of Automatic Control
Cape Town, South Africa. August 24-29, 2014

Trajectory Tracking of Batch Product Quality Using Latent Variable Models ⋆

Eduardo B. Lopez-Montero * **Ognjen Marjanovic** ** **Barry Lennox** ***

* *School of Electrical and Electronic Eng., The University of Manchester, UK (e-mail: Eduardo.LopezMontero@postgrad.manchester.ac.uk)*
** *School of Electrical and Electronic Eng., The University of Manchester, UK (e-mail: Ognjen.Marjanovic@manchester.ac.uk)*
*** *School of Electrical and Electronic Eng., The University of Manchester, UK (e-mail: Barry.Lennox@manchester.ac.uk)*

Abstract: A practical strategy for controlling batch product quality evolution by means of latent variable models and intermittent measurements is presented. The methodology is based on the identification of data-based models using multivariate statistical methods such as Partial Least Squares (PLS). PLS is able to identify models with a reduced number of latent variables, which account for most of the process variability. The data-based models are employed along with a moving window strategy in order to predict product quality throughout the batch operating time. The predictions can be utilized within a Model Predictive Control (MPC) architecture so that trajectory tracking control can be directly applied to batch product quality. A simulated example of fed-batch aerobic growth of Saccharomyces Cerevisiae is used to demonstrate the capabilities of the proposed trajectory tracking controller.

Keywords: Batch processes control; Partial least squares; Model Predictive Control.

1. INTRODUCTION

Batch and semi-batch processes constitute a very important part of the chemical industry. Furthermore, new market environment has generated an increase in the demand of low-volume high added-value products (Barbosa-Póvoa (2007)). And since it has been shown that the development and application of control strategies for chemical processes can improve profitability and reduce operational costs without the need of plant redesign (Marlin et al. (1991)), an increased interest has arise among researchers and industry for developing new batch control techniques. In batch processes the main objective is the achievement of a specific product quality by the end of the batch operating time, that commonly corresponds to a non-equilibrium point. This represents an interesting problem for the control community, specially due to the typical characteristics exhibited by these type of processes (Bonvin et al. (2006)): complex physical phenomena, irreversible time-variant and non-linear dynamics, and lack of quality sensors. In the past, batch product quality was controlled by implementing predetermined input set-point trajectories, which were optimized off-line, in conjunction with simple controller algorithms like time-optimal control or standard Proportional-Integral-Derivative (PID). But the use of these control strategies made the batch product quality susceptible to disturbances and changes in initial conditions.

In recent years it has been found that model based controllers offer improved performance over simpler control algorithms (Aziz et al. (2000)). Model Predictive Control (MPC), which utilizes a plant model within its formulation, has been the base for many batch process controllers. MPC has been widely accepted in industry and successfully used in real applications (Jämsä (2007)). But reliable models are not always available, and developing rigorous models for batch processes could be very time consuming, requiring a deep theoretical understanding of the process. Recently, multivariate statistical methods have been increasingly used to identify data-based models in order to control and monitor batch processes. These methods offer considerable advantages over rigorous ones: They do not require detailed theoretical knowledge of the process, models are relatively easy to build and keep up to date. Amongst the data-driven methodologies used to identify batch process models Principal Component Analysis (PCA) and Partial Least Squares (PLS), which were proposed by MacGregor and co-workers (Kresta et al. (1991); MacGregor and Kourti (1995)), have received particular attention form researchers.

In Golshan et al. (2011) a latent variable MPC is presented for trajectory tracking regulation, the methodology is based on PCA models for capturing the relationship among process variables in order to perform estimations of future variable values. For this control approach the target trajectories are usually the ones belonging to some

⋆ This research project is funded by the Mexican National Council for Science and Technology (CONACyT), grant number: 172607/199502; and the Mexican Secretariat of Public Education (SEP), grant number: BC-1073.

key process variables rather than the actual values of desired batch quality. Hence, it is assumed that the desired product quality will be achieved if some secondary process variables follow their predetermined trajectories. However, such assumption is not always true, specially in processes where changes in initial conditions are common (Russell et al. (1998)).

A batch control focused on end-point quality was proposed by Flores-Cerrillo and MacGregor (2004), which is based on a PLS model that relates the values of the process variables to batch product quality. This control strategy does not consider product quality measurements in a new batch run. Consequently, this strategy is likely to be affected by disturbances or changes in batch condition parameters (e.g. change in raw materials characteristics). Some work have been made to overcome this issue, Yacoub and MacGregor (2011) proposed to identify different models for each disturbance case-scenario, which requires the availability of data belonging to the most frequent type of disturbances. Also, the inclusion of a disturbance model to offset future measured variables predictions have been proposed by Wan et al. (2012), but as in the trajectory tracking approach proposed by Golshan et al. (2011), assuming that by controlling and estimating accurately some key variables will result in reaching desired tight product specifications is not generally true. Therefore, a practical data-based MPC strategy for trajectory tracking of batch product quality is proposed in this paper, which represents a feasible alternative to tackle the issues mentioned above. Even though it is not possible to continuously measure product quality, it can often be quantified intermittently through laboratory assays. Using the intermittent measured data, a PLS model can be constructed employing the procedure developed by Marjanovic et al. (2006). Then, the PLS model can be used to make predictions of batch product quality. Such predictions are based on the PLS model and a moving window approach. The predictions can be used within a MPC formulation for trajectory tracking control of product quality.

The remainder of the paper is organized as follows: Section 2 details the PLS model building procedure, and the proposed trajectory tracking formulation for controlling batch product quality. In Section 3 a structured model for the aerobic growth of Saccharomyces cerevisiae is used as a simulation benchmark, in order to demonstrate the proposed controller capabilities; finally some concluding remarks are included in Section 4.

2. TRAJECTORY TRACKING WITH INTERMITTENT MEASUREMENTS

The proposed trajectory tracking controller for batch product quality can be divided into four stages. The first one consists in the identification of a PLS model using intermittent measurements. Secondly, predictions of future product quality are made using the identified PLS model and a moving window strategy. The third stage consists in incorporating the product quality predictions in a MPC architecture so as to perform trajectory tracking control. The final step consists in employing, if available, intermittent measurements during a new batch run in order to reject disturbances. These four stages are described in the next subsections.

2.1 Model identification

In order to identify the PLS model, process variables are split into inputs (readily measured process variables) and outputs (batch product quality). The procedure employed to identify the PLS model is the one presented by Marjanovic et al. (2006). Such method takes into consideration intermittent measurements of batch product quality, and a new arrangement of the data. A Pseudo-batch is created for each of the intermittent measurements, and then they are aligned toward their end-points. As a result, a short window PLS model is identified. Fig. 1 is included to illustrate the new data alignment. In such figure, data from two different batch runs is depicted. Three measurements of product quality were made during each batch. As a result, six pseudo-batches were formed and aligned toward their end-points, as illustrated in Fig. 1. After this, a modelling window is chosen. Such modelling window is equal to the smallest pseudo-batch. For PLS model identification, the intermittent measurements are considered as the outputs, while the rest of the process variables are considered as inputs. For more details on the data arrangement refer to the cited paper.

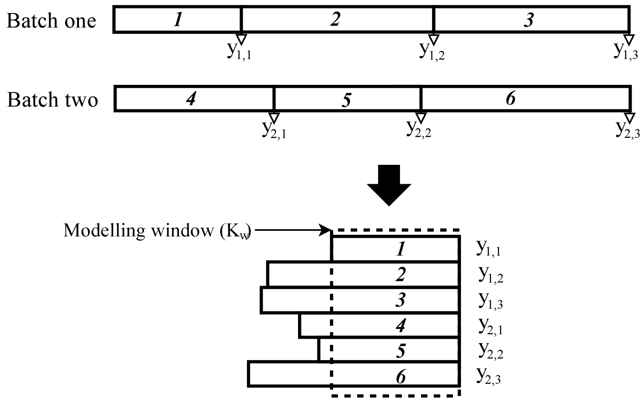

Fig. 1. Pseudo-batch wise unfolding

The new data arrangement has I_w rows (pseudo-batches) and is formed by two matrices: X_w (input) and Y_w (output), where $X_w \in \mathbb{R}^{I_w \times JK_w}$ and $Y_w \in \mathbb{R}^{I_w \times n_y}$, J represents the number of input variables, n_y the number of outputs, and K_w denotes the modelling window size or number of sample instants used to build the new data matrices. The PLS model is given by:

$$X_w = TP^T + E \tag{1}$$

$$Y_w = UQ^T + F \tag{2}$$

where $T \in \mathbb{R}^{I \times n_{lv}}$, $P \in \mathbb{R}^{JK_w \times n_{lv}}$, and $E \in \mathbb{R}^{I \times JK_w}$ are the input scores, loadings and residuals matrices respectively. Similarly, $U \in \mathbb{R}^{I \times n_{lv}}$, $Q \in \mathbb{R}^{n_y \times n_{lv}}$, and $F \in \mathbb{R}^{I \times n_y}$ are the output scores, loadings and residuals matrices, respectively. The number of latent variables retained by the model is denoted as n_{lv}, which is commonly chosen through cross-validation (MacGregor and Kourti (1995)). The input and output scores are related by a diagonal matrix $B \in \mathbb{R}^{n_{lv} \times n_{lv}}$ so that $U = TB$. The non-linear iterative partial least squares (NIPALS) regression algorithm [1] (Kresta et al., 1991) is normally used to obtain

[1] NIPALS is applied to scaled data (zero mean and unit variance).

the PLS model. In this algorithm an additional weighting matrix $W \in \mathbb{R}^{JK \times n_{lv}}$ is used to calculate the input scores: $T = X_w W \left(P^T W\right)^{-1}$. Consequently, the PLS model can be expressed as follows:

$$Y_w = X_w \underbrace{W \left(P^T W\right)^{-1} B Q^T}_{\Theta} + F^* \tag{3}$$

where F^* represents the residuals matrix formed by negligible information if an appropriate n_{lv} was chosen. After discarding F^*, the PLS model becomes: $Y_w = X_w \Theta$.

2.2 Moving window estimation

After identifying the PLS model using stored data, it can then be employed for the prediction of product quality during new batch runs. Such prediction is based on a moving window approach. The proposed moving window approach can be described as follows: Consider that the current time instant of a new batch run is k. The first step consists in building an input vector z_k. Such vector is formed by the $K_w - 1$ past samples of the input variables:

$$z_k = \left[x^T_{me|k-K_w+2 \to k} \;\; u^T_{mv|k-K_w+2 \to k}\right]^T \tag{4}$$

$$x_{me} = \left[x_1 \cdots x_{n_x}\right]^T \in \mathbb{R}^{n_x \times 1} \tag{5}$$

$$u_{mv} = \left[u_1 \cdots u_{n_u}\right]^T \in \mathbb{R}^{n_u \times 1} \tag{6}$$

where x_{me} and u_{mv} are the vectors formed by the readily measured and manipulated variables, respectively; n_x and n_u represent the number of readily measured variables and the number of manipulated variables, respectively. Assuming that future manipulated variables $u_{mv|k+1}$ are available after solving a minimization problem, then future values of the readily measured process variables $x_{me|k+1}$ can be estimated by means of the PLS model and missing data methods. Algorithms for inferring missing data can be found in Nelson et al. (1996). Where the underlying data pattern is used to calculate missing variable values using the known variable values. Using the Projection to the Model Plane (PMP) method, input variables $\mathbf{x}$ are split as follows: $\mathbf{x}^T = [\mathbf{x}^{*^T} \;\; \mathbf{x}^{\sharp^T}]$, where $\mathbf{x}^*$ refers to the known values and $\mathbf{x}^\sharp$ to the missing or unknown values. Also, loading matrix P from (1), can be split in P^* and $P^\sharp$, corresponding to $\mathbf{x}^*$ and $\mathbf{x}^\sharp$, respectively. Afterwards, the missing variable values can be deduced using an optimal score vector $\hat{\mathbf{t}}$. Such score vector is the result of solving the next minimization problem:

$$J_{\mathbf{t}} = \frac{1}{2}(\mathbf{x}^* - P^*\mathbf{t})^T(\mathbf{x}^* - P^*\mathbf{t}). \tag{7}$$

The optimal solution ($\hat{\mathbf{t}}$) is obtained by taking the derivative of (7) and setting the result equal to zero, $\frac{\partial J_{\mathbf{t}}}{\partial \mathbf{t}} = 0$:

$$\hat{\mathbf{t}} = (P^{*^T} P^*)^{-1} P^{*^T} \mathbf{x}^* \tag{8}$$

Then, the missing variable values can be deduced straightforwardly: $\mathbf{x}^\sharp = \mathbf{P}^\sharp \hat{\mathbf{t}}$. Therefore, values of the readily measured process variables at sample instant $k+1$ (denoted as $\widehat{x}_{me|k+1}$) can be expressed as a function of past $K_w - 1$ input variables (z_k), and the future manipulated variables ($u_{mv|k+1}$):

$$\widehat{x}_{me|k+1} = \underbrace{P^\sharp \left(P^{*T} P^*\right)^{-1} P^{*T} \left[z_k^T \;\; u^T_{mv|k+1}\right]^T}_{f_1\left(z_k, u_{mv|k+1}\right)} \tag{9}$$

A similar approach is applied for the prediction of the future quality variables, denoted as $\widehat{y}_{k+1}$. From (3) and (9), the future predicted output variables can be expressed as a function of past $K_w - 1$ input variables (z_k), and the future manipulated variables ($u_{mv|k+1}$):

$$\widehat{y}_{k+1} = \underbrace{\Theta^T \begin{bmatrix} z_k \\ \widehat{x}_{me|k+1} \\ u_{mv|k+1} \end{bmatrix}}_{f_2\left(z_k, u_{mv|k+1}\right)} \tag{10}$$

After $\widehat{x}_{me|k+1}$ and $\widehat{y}_{k+1}$, have been estimated using (9) and (10), the modelling window is moved one sample instant forward. Then, the measured variables and the product quality values at time instant $k+2$ can be computed using $u_{mv|k+2}$ and the formerly estimated value of $\widehat{x}_{me|k+1}$. This prediction procedure is repeated recursively up to the batch end-point (denoted as K).

2.3 Trajectory tracking control

The trajectory tracking control is carried out in a shrinking horizon manner. The future manipulated variable trajectories are optimized at each control decision point (k_p). The objective is to calculate the values of the manipulated variables that minimize $\widehat{y} - \overline{y}$. Where $\widehat{y}$ and $\overline{y}$ denote the predicted quality, and the desired quality, respectively. The optimal future manipulated variable trajectories are applied to the batch process. Then, the optimal manipulated variable trajectories are calculated again at the next decision point. This procedure is repeated until the end of the batch (K). Assuming that the current control decision point is k, the predicted future quality trajectory ($\widehat{y}_{k+1 \to K}$) can be computed employing the approach presented in the previous subsection. Therefore, the optimization of the future manipulated variable trajectories can be formulated as follows:

$$\begin{aligned} \underbrace{\min}_{u_{mv|k=k_p}} \quad & \left|\left|\widehat{y}_{k+1\to K} - \overline{y}_{k+1\to K}\right|\right|^2_{Q_1} + \\ & \left|\left|u_{mv|k+1\to k+M} - u_{mv|k\to k+M-1}\right|\right|^2_{Q_2} \\ \textbf{s.t.} & \\ \widehat{x}_{me|k+1} &= f_1\left(z_k, u_{mv|k+1}\right) \\ \widehat{y}_{k+1} &= f_2\left(z_k, u_{mv|k+1}\right) \\ \Delta u &\le \Delta u_{max} \\ U_{lb} &\le u_{mv} \le U_{ub} \end{aligned} \tag{11}$$

where $Q_1 \in \mathbb{R}^{n_y \cdot (K-k) \times n_y \cdot (K-k)}$ and $Q_2 \in \mathbb{R}^{M \cdot n_u \times M \cdot n_u}$ are the symmetric and positive definite weighting matrices for trajectory error ($\widehat{y} - \overline{y}$) and control change rate (Δu), respectively. M is the control horizon, Δu_{max} is the maximum control change rate allowed, $U_{lb} \in \mathbb{R}^{n_u \times 1}$ and $U_{ub} \in \mathbb{R}^{n_u \times 1}$ are the vectors formed by the lower (lb) and upper (ub) bounds of the manipulated variables.

2.4 Disturbance rejection with intermittent measurements

The target quality trajectory can be adjusted using the difference between the estimated and the measured values of batch quality. Such difference or offset can be computed when intermittent measurements are available. Then, the offset can be used for rejecting un-modelled disturbances. The method for calculating the offset, denoted as Δy, is described next: Suppose that a new intermittent measurement, denoted as y_{k_s}, is available at sample instant k_s. Then, using the PLS model depicted in (3) and the past K_w values of both x_{me} and u_{mv}, the estimated output $(\widehat{y}_{k_s})$ can be obtained as follows:

$$\widehat{y}_{k_s}^T = \left[x_{me|k_s-K_w+1\rightarrow k_s}^T \ u_{mv|k_s-K_w+1\rightarrow k_s}^T \right] \Theta \quad (12)$$

then, the offset can be defined as:

$$\Delta y = \widehat{y}_{k_s} - y_{k_s} \quad (13)$$

Therefore, the new target trajectory is:

$$\overline{y}_{new} = \overline{y}_{old} + \mathbf{1} \cdot \Delta y^T \quad (14)$$

where $\mathbf{1}$ represents a $K \times 1$ column vector of ones. Notice that if there are no new intermittent measurements during a new batch run, then $\overline{y}_{new} = \overline{y}_{old}$. The new target quality trajectory is then used in the optimization problem depicted in (11). If an un-modelled disturbance is presented and an intermittent measurement has been taken, the control algorithm will be able to react against such disturbance by means of the offset Δy. An example of this is included in the next section.

3. CASE STUDIES

A simulation benchmark developed by Lei et al. (2001) was employed to asses the performance of the proposed controller. The benchmark corresponds to a biochemically structured model for growth of Saccharomyces cerevisiae. The process variables are: the concentration of glucose, pyruvate, acetaldehyde, acetate and ethanol; the active cell material; the acetaldehyde dehydrogenase proportional to the measured activity; the specific oxygen uptake rate (OUR); the specific carbon dioxide evolution rate (qCO2); and the volume. During the simulations only the OUR, qCO2 and volume were considered to be continuously measured in order to agree with industrial practice. The manipulated variable is the glucose feed rate, and the quality related variable (output) is the Biomass concentration.

3.1 Data collection and model identification

Data from 30 batches was gathered in order to identify a PLS model. Each single batch had a duration of $K = 21.5$ hours. The process variables were measured every 0.1 hours, and a filtered pseudo random binary signal (PRBS) was added on top of the nominal substrate feed rate of 21 g/hr, in order to excite process dynamics. It was assumed that a few biomass concentration measurements were taken during each batch run. Samples were taken at the end of each simulated batch, and around the 6^{th}, 12^{th} and 18^{th} hours (± 0.5 hours). A total of four measurements of product quality were taken during each batch run, therefore 120 pseudo-batches were created and aligned following the procedure depicted in Fig. 1. The length for the modelling window was selected to be $K_w = 5$ hours (50 samples). Afterwards, a PLS model was identified following the methodology described in Section 2. The number of latent variables retained for the model was chosen to be 4 through leave-one-out cross-validation (Diana and Tommasi, 2002).

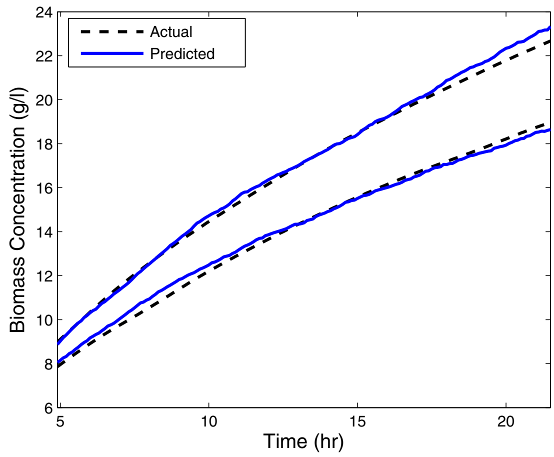

Fig. 2. Predicted values of Biomass concentration

Results from the simulation of two batch runs are presented in Fig. 2 in order to demonstrate the prediction capabilities of the PLS model. During the simulations a PRBS was appended to the nominal substrate feed rate, and a white noise signal with a Signal to Noise Ratio [2] (SNR) = 40 dB was added to the measurements. In Fig. 2 it can be seen that the predicted trajectories of Biomass concentration are close to the actual ones. The results displayed in Fig. 3 correspond to the estimation of Carbon dioxide during two simulated batch runs, where noise and PRBS were added to the measurements and the substrate feed rate, respectively. In Fig. 3 the black dashed line corresponds to the noisy (measured) values, the red solid line corresponds to the actual values and the blue solid line represents the estimated values of Carbon dioxide. It is noticeable that the PLS model is able to provide accurate estimations of readily measured process variables, even if the measurements used to make such estimations are affected by noise.

3.2 Trajectory tracking of batch product quality

The objective of the MPC control formulation consists in tracking a desired product quality evolution throughout the batch operating time. Three case scenarios were considered in order to evaluate the performance of the proposed controller. The following parameters were utilised during the simulations: $k_p = \{K_w,\ K_w+1,\ K_w+2,\ \cdots,\ K-1\}$; $lb = 10\ g/hr$; $ub = 30\ g/hr$; $M = 3$; and $\Delta u_{max} = 10\ g/hr$. For the first case scenario, the controller objective consisted on tracking a nominal Biomass trajectory; for the second case scenario,

[2] $\mathrm{SNR} = 20 \log \left(\frac{\mathrm{RMS}_{signal}}{\mathrm{RMS}_{noise}} \right) dB$

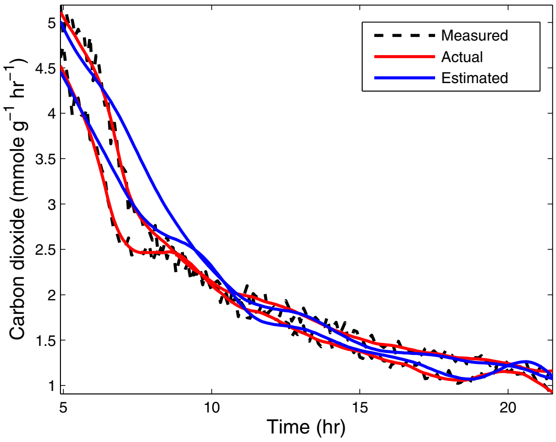

Fig. 3. Carbon dioxide evolution rate

a modified Biomass trajectory ($+1g/l$) was considered. For these two scenarios white noise with a SNR $= 40\ dB$ was added to all measurements, and no intermittent measurements were considered to be taken during the simulations. Results from the first two cases are shown in Fig. 4 and Fig. 5. The first 21.5 hours of Fig. 4 correspond to the first simulated batch, where no un-modelled disturbances where considered. As would be expected, the results obtained in open-loop using a nominal substrate feed rate and the ones using the proposed controller were satisfactory, as the desired Biomass concentration trajectory was closely tracked. However, the outcome can be used to demonstrate that the control sequence obtained with the proposed approach is close to the nominal substrate feed rate even though measurements were affected by noise. This can be noted by inspecting the first 21.5 hours of Fig. 5, which correspond to the manipulated variable trajectory of the first simulated batch.

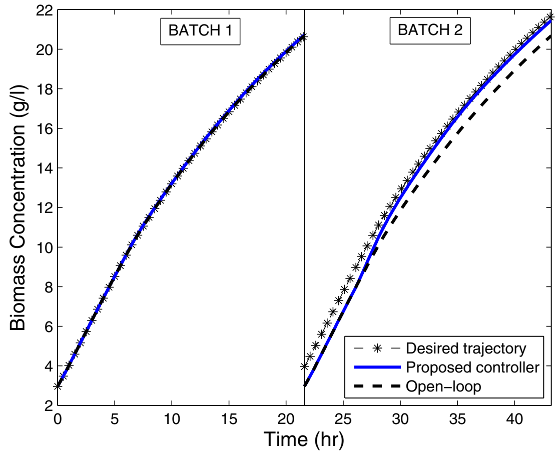

Fig. 4. Tracking batch product quality set point

The last 21.5 hours of Fig. 4 and Fig. 5 correspond to the second scenario were a modified desired trajectory was considered. In Fig. 4 it can be noted that the target trajectory was closely followed only by the proposed controller, this was possible by adding an offset ($\Delta y = 1$) to the desired quality trajectory when solving the minimization

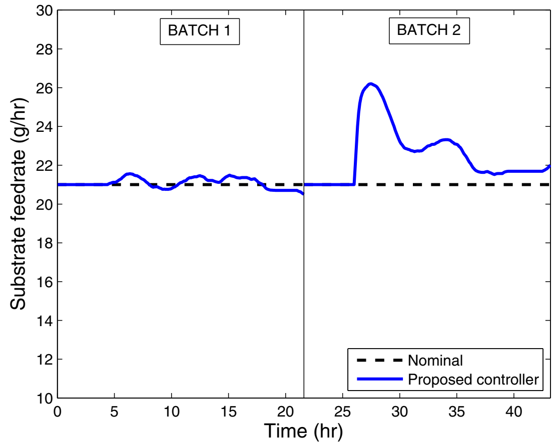

Fig. 5. The computed manipulated variable trajectory

problem depicted in (11). The adjustments made by the controller to the manipulated variable can be seen from the 25^{th} to the 43^{th} hour in Fig. 5, these adjustments to the substrate feed rate were made without violating any of the constraints (i.e. $10 \leq u_{mv} \leq 30$ and $\Delta u \leq 10$).

The third case scenario consisted on a batch run subject to an un-modelled disturbance. The disturbance was selected to be a change in the substrate concentration. The change occurred around the 5^{th} hour. Such change consisted on a 8% decrease of the substrate concentration (i.e. from a nominal 100 g/l to 92 g/l). For this last simulated scenario, intermittent measurements were assumed to be taken at the 9^{th} and 17^{th} hour. Additionally, the proposed control approach was compared to the batch end-point controller addressed in Wan et al. (2012). The latter was done in order to show the advantages of taking into consideration intermittent product quality measurements. Results of the third simulated scenario are shown in Fig. 6 and Fig. 7, which correspond to the product quality and manipulated variable trajectories, respectively. By a close inspection to Fig. 6 and Fig. 7 it can be noted that the proposed controller and the end-point controller increase the substrate feed rate to cope with the disturbance. However, the proposed controller achieved a better tracking performance than the batch end-point controller. Table 1 is included to provide a quantitative comparison of the results obtained with the proposed controller, the end-point controller and open-loop operation, the performance index used for such comparison was the Root Mean Square Error (RMSE) of the Biomass trajectory.

Table 1. Trajectory tracking results

Control methods	RMSE
Proposed	0.155
End-point	0.421
Open-loop	0.667

By inspecting table 1 it can bee noted that the proposed control approach provides the best results compared to the end-point controller and open-loop operation, which confirms the results displayed in Fig. 6. The control approach proposed by Wan et al. (2012), reacted against the perturbation due to a disturbance model within its formulation, which is used to calculate a difference between

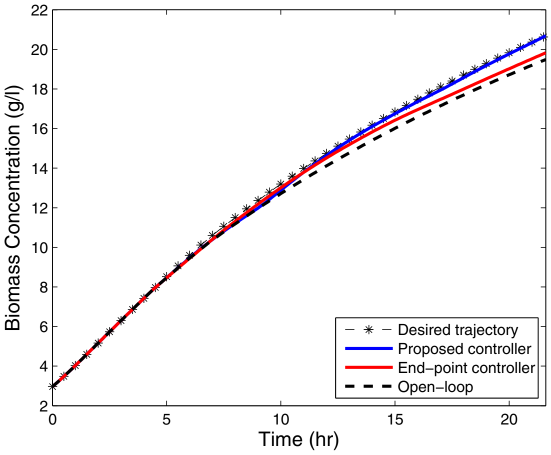

Fig. 6. Tracking product quality in case of disturbances

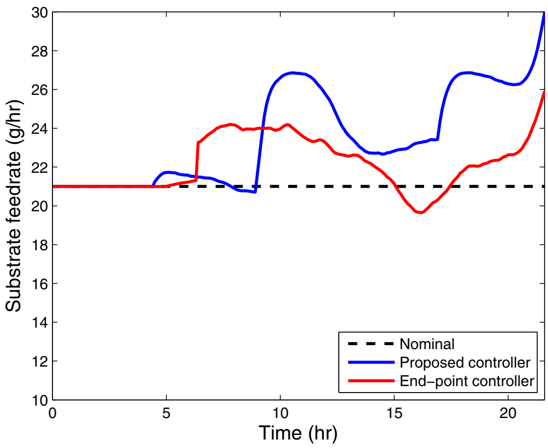

Fig. 7. Manipulated variable trajectory

estimated and the actual values of the readily measured variables. This strategy is based on the ability of such process variables to convey desirable end-quality properties if they follow a predetermined behaviour, which is not always the case. The smallest RMSE, indicating a closer trajectory tracking of the desired product quality, was obtained using the proposed controller. Such performance was the result of detecting and rejecting the un-modelled disturbance by calculating Δy, as described in Section 2.4.

4. CONCLUSIONS

A control methodology for tracking batch product quality has been proposed. The controller is based on intermittent measurements and a latent variable PLS model identified using stored data. Such model is able to describe a batch process without being affected by the high correlation existing among the variables. Incorporating the model and a moving window strategy to a MPC framework, gives the possibility to carry out trajectory tracking of batch product quality. By means of a simulation benchmark it was demonstrated that the controller was able to achieve quality specifications, even in presence of disturbances, changes in nominal conditions and considering noisy measurements. The proposed controller can reject disturbances within a batch run due to its capability of using intermittent measurements in its formulation. Reason why it can perform better than conventional end-point controllers and therefore, is a suitable option for batch process control.

REFERENCES

Aziz, N., Hussain, M.A., and Mujtaba, I.M. (2000). Performance of different types of controllers in tracking optimal temperature profiles in batch reactors. *Computers & Chemical Engineering*, 24, 1069–1075.

Barbosa-Póvoa, A.P. (2007). A critical review on the design and retrofit of batch plants. *Computers & Chemical Engineering*, 31(7), 833–855.

Bonvin, D., Srinivasan, B., and Hunkeler, D. (2006). Control and Optimization of Batch Processes - Improvement of process operation in the production of specialty chemicals. *IEEE Control Systems Magazine*, (Dec.), 34–45.

Diana, G. and Tommasi, C. (2002). Cross-validation methods in principal component analysis: A comparison. *Statistical Methods & Applications*, 11(1), 71–82.

Flores-Cerrillo, J. and MacGregor, J.F. (2004). Control of batch product quality by trajectory manipulation using latent variable models. *Journal of Process Control*, 14(5), 539–553.

Golshan, M., MacGregor, J.F., and Mhaskar, P. (2011). Latent variable model predictive control for trajectory tracking in batch processes: Alternative modeling approaches. *Journal of Process Control*, 21(9), 1345–1358.

Jämsä, S.L. (2007). Future trends in process automation. *Annual Reviews in Control*, 31(2), 211–220.

Kresta, J.V., MacGregor, J.F., and Marlin, T.E. (1991). Multivariate statistical monitoring of process operating performance. *The Canadian Journal of Chemical Engineerging*, 69, 35–47.

Lei, F., Rotbø ll, M., and Jø rgensen, S.B. (2001). A biochemically structured model for Saccharomyces cerevisiae. *Journal of biotechnology*, 88(3), 205–21.

MacGregor, J.F. and Kourti, T. (1995). Statistical process control of multivariate processes. *Control Engineering Practice*, 3(3), 403–414.

Marjanovic, O., Lennox, B., Sandoz, D., Smith, K., and Crofts, M. (2006). Real-time monitoring of an industrial batch process. *Computers & Chemical Engineering*, 30(10-12), 1476–1481.

Marlin, T.E., Perkins, J.D., Barton, G.W., and Brisk, M.L. (1991). Benefits from process control: results of a joint industry-university study. *Journal of Process Control*, 1, 68–83.

Nelson, P.R.C., Taylor, P.A., and MacGregor, J.F. (1996). Missing data methods in PCA and PLS : Score calculations with incomplete observations. *Chemometrics and Intelligent Laboratory Systems*, 35, 45–65.

Russell, S.A., Kesavan, P., Lee, J.H., and Ogunnaike, B.A. (1998). Recursive data-based prediction and control of batch product quality. *AIChE Journal*, 44(11), 2442–2458.

Wan, J., Marjanovic, O., and Lennox, B. (2012). Disturbance rejection for the control of batch end-product quality using latent variable models. *Journal of Process Control*, 22(3), 643–652.

Yacoub, F. and MacGregor, J.F. (2011). Robust Processes Through Latent Variable Modeling and Optimization. *AIChE Journal*, 57(5), 1278–1287.